Drift Chamber with Cluster Counting for CEPC

Guang Zhao, Linghui Wu, Mingyi Dong, Gang Li, Zhefei Tian, Zhenyu Zhang, Xu Gao, Shuaiyi Liu, Shengsen Sun

zhaog@ihep.ac.cn

Jan 30th, 2024

1st DRD1 Collaboration Meeting

中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

WUHAN UNIVERSITY 1893 武汉大学

JILIN UNIVERSITY·CHINA 1946 吉林大学

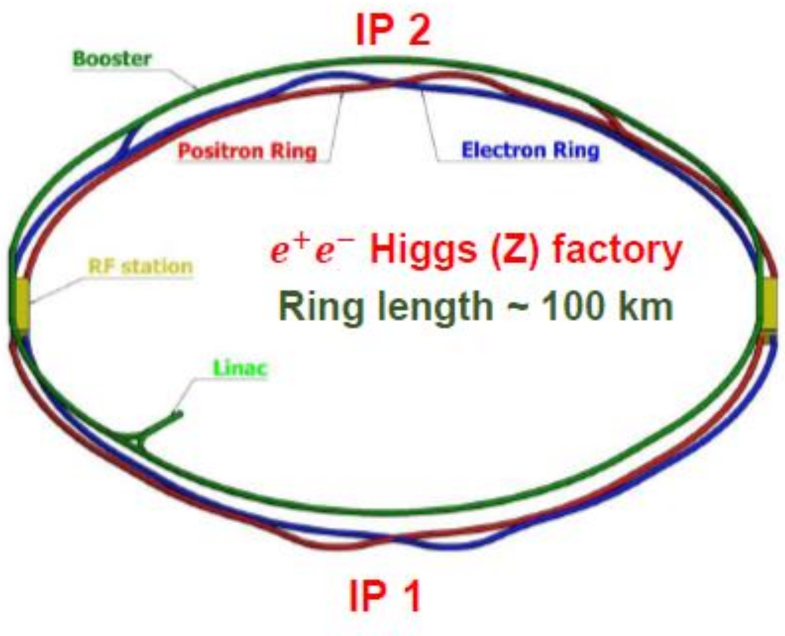
INFN
Istituto Nazionale di Fisica Nucleare

Outline

- Introduction
- Detector optimization
- Reconstruction algorithm for CC
- Prototype experiments
- Summary

The Circular Electron Positron Collider (CEPC)

- The CEPC was proposed in 2012 right after the Higgs discovery. It aims to start operation in 2030s, as an e^+e^- Higgs / Z Factory.
- To produce Higgs / W / Z / top for high precision Higgs, EW measurements, studies of flavor physics & QCD, and probes of physics BSM.
- It is possible to upgrade to a pp collider (SppC) of $\sqrt{s}$ ~ 100 TeV in the future.



Particle	$E_{c.m.}$ (GeV)	Years	SR Power (MW)	Lumi. /IP ($10^{34}cm^{-2}s^{-1}$)	Integrated Lumi. /yr (ab^{-1}, 2 IPs)	Total Integrated L (ab^{-1}, 2 IPs)	Total no. of events
H*	240	10	50	8.3	2.2	21.6	4.3×10^6
			30	5	1.3	13	2.6×10^6
Z	91	2	50	192**	50	100	4.1×10^{12}
			30	115**	30	60	2.5×10^{12}
W	160	1	50	26.7	6.9	6.9	2.1×10^8
			30	16	4.2	4.2	1.3×10^8
$t\bar{t}$	360	5	50	0.8	0.2	1.0	0.6×10^6
			30	0.5	0.13	0.65	0.4×10^6

* Higgs is the top priority. The CEPC will commence its operation with a focus on Higgs.
** Detector solenoid field is 2 Tesla during Z operation, 3Tesla for all other energies.
*** Calculated using 3,600 hours per year for data collection.

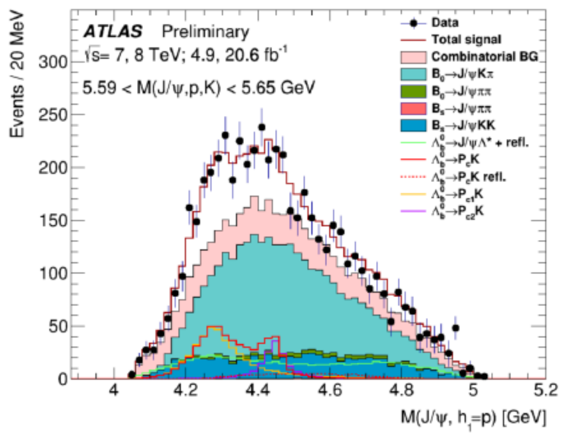
Particle identification

- **PID is essential for CEPC, especially for flavor physics**
  - Suppressing combinatorics
  - Distinguishing between same topology final-states
  - Adding valuable additional information for flavor tagging of jets
  - …

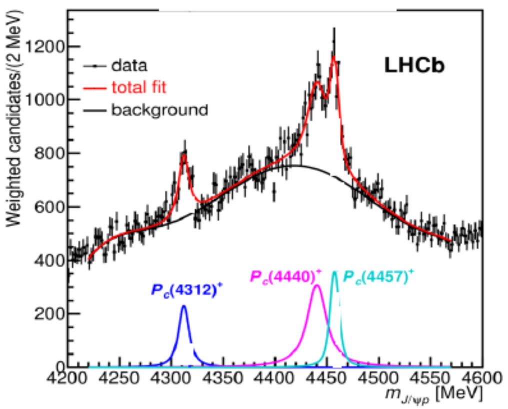
Pentaquark search in $\Lambda_b \to J/\Psi pK$

Without PID

[ATLAS-CONF-2019-048]

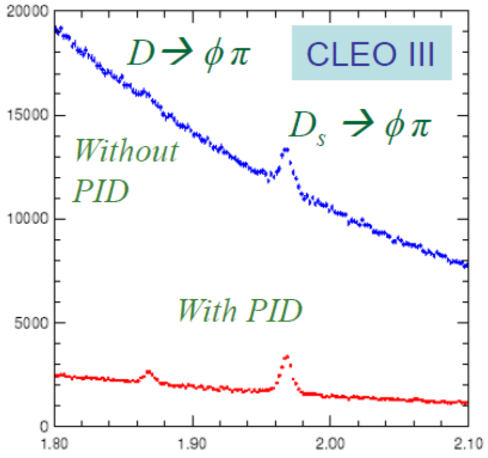


With PID ~ no background



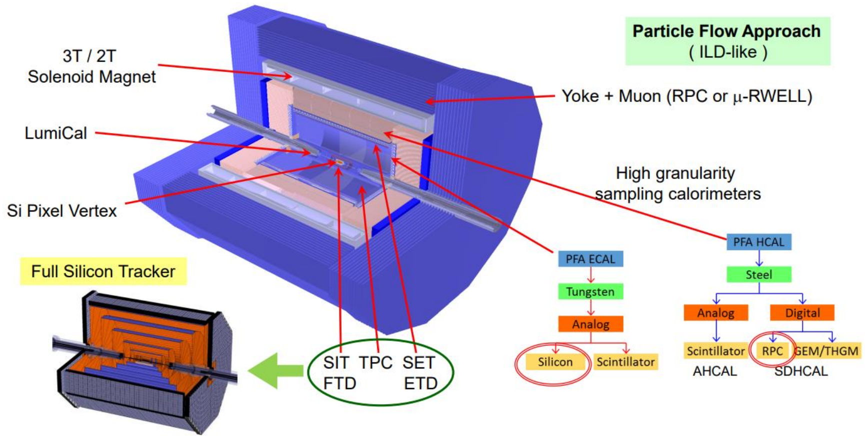
[LHCb, PRL 122 (2019) 222001]

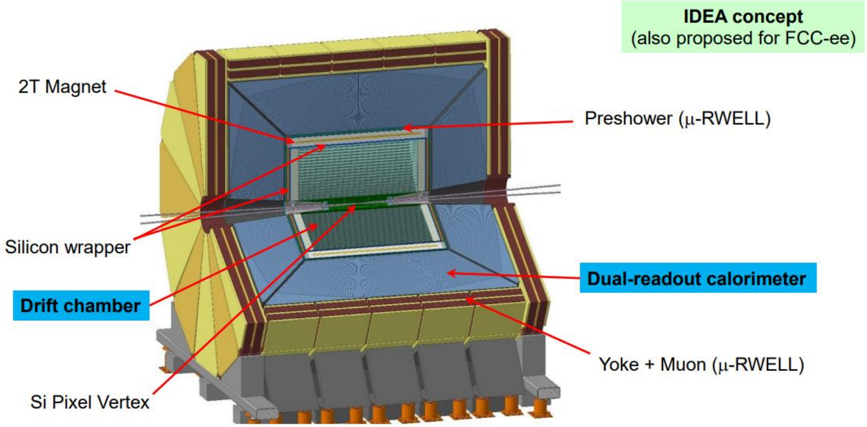
$D_{(s)} \to \phi\pi$



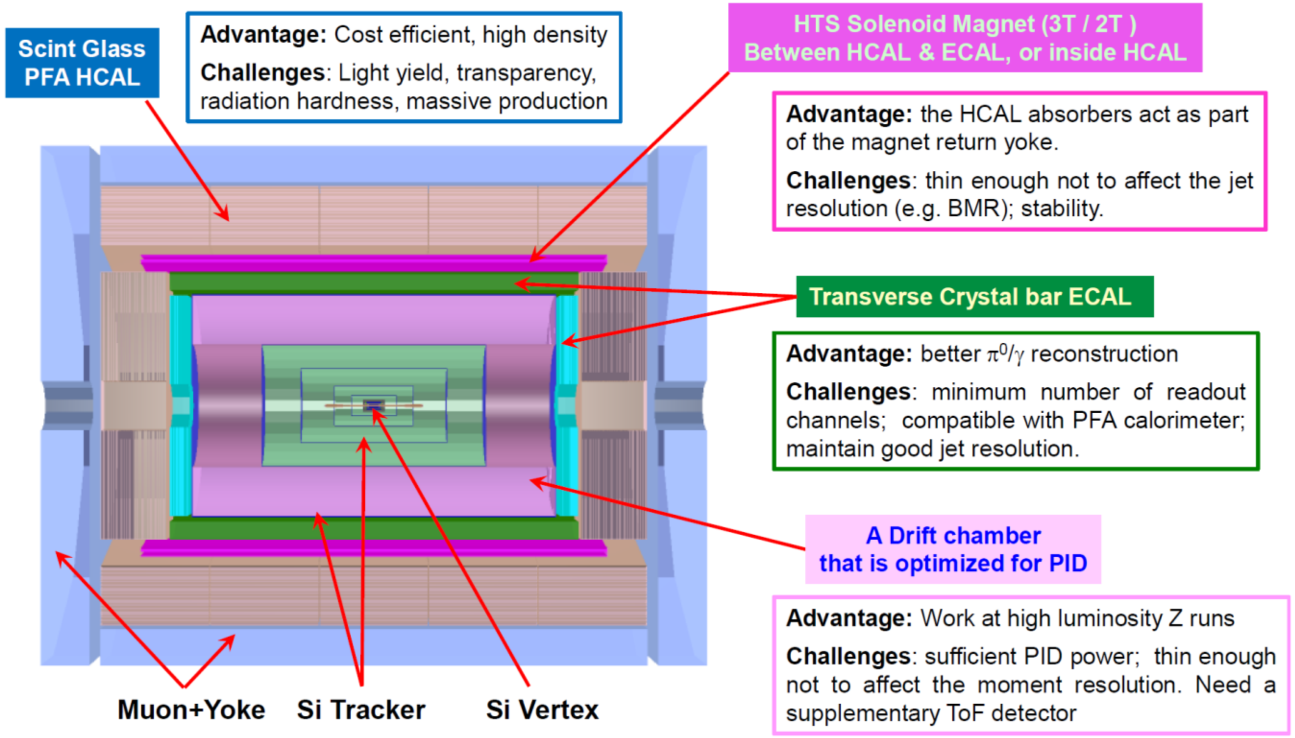
* Figures from Guy Wilkinson's slides

CEPC detector concepts



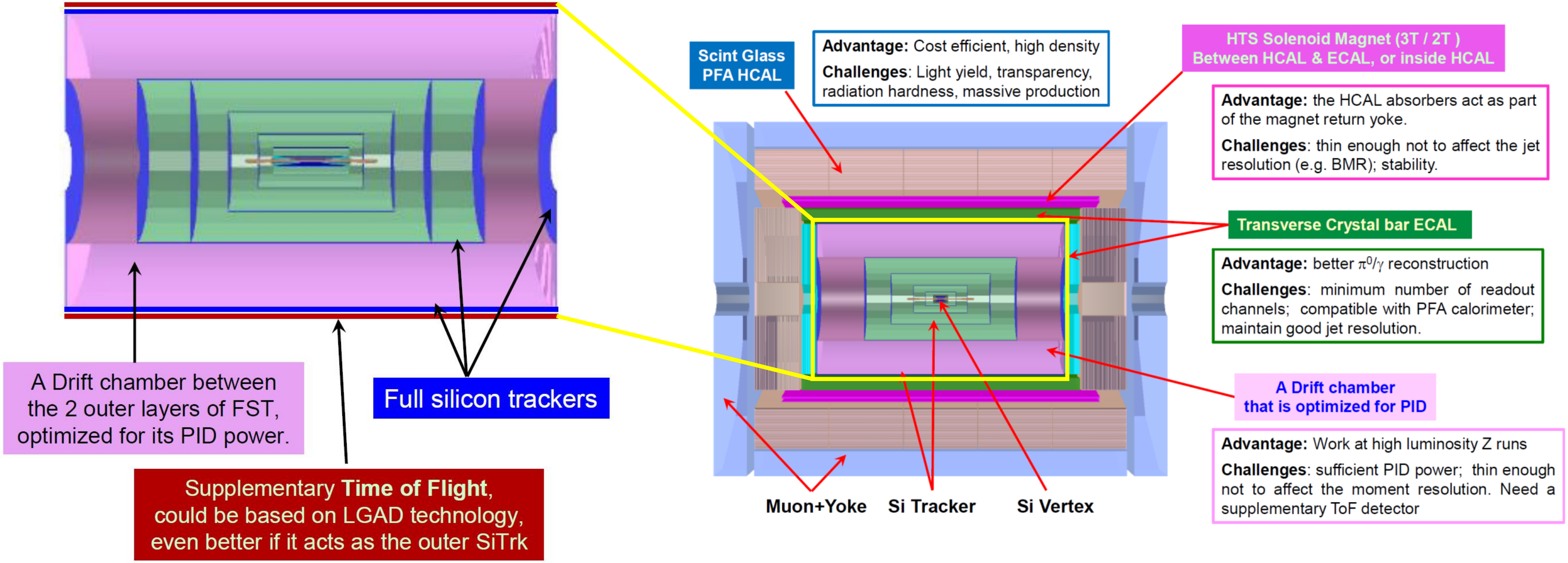


CEPC 4th concept



➔ See yesterday's talk from Nikola for IDEA DC

CEPC 4th concept detector

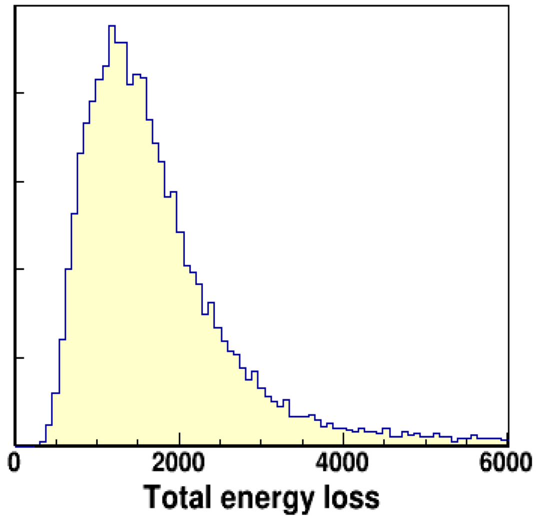


Preliminary PID requirement: >2σ K/π separation for 20 GeV/c tracks

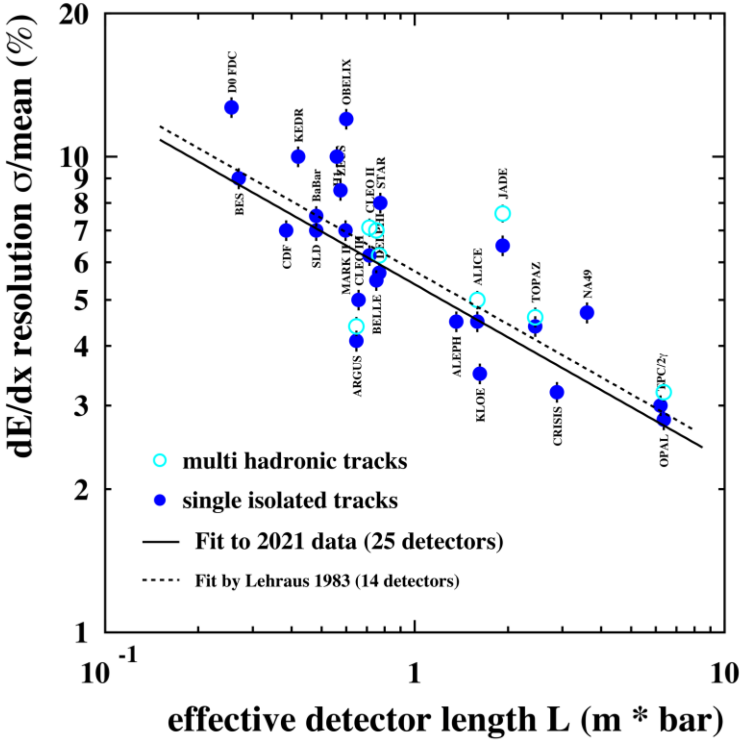
Energy loss measurement: dE/dx

- **Main mechanism: Ionization of charged tracks**
- **Traditional method: Total energy loss (dE/dx)**
  - Landau distribution due to secondary ionizations
  - Large fluctuation from many sources: energy loss, amplification …

Integrated charge



"Lehraus" Plot 2021*



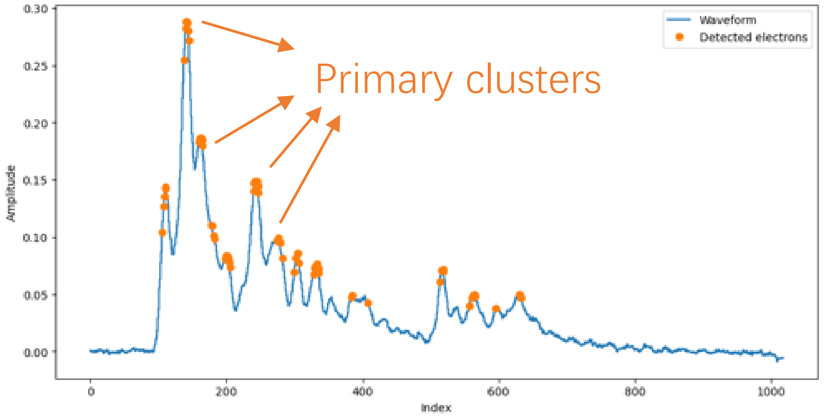
- **Fit by Lehraus 1983:**
  - dE/dx res. = **5.7** $* L^{-0.37}$ (%)
- **Fit in 2021:**
  - dE/dx res. = **5.4** $* L^{-0.37}$ (%)
- **No significant improvement in the past 40 years**

* From Michael Hauschild's talk @ RD51 workshop

Cluster counting measurement (CC)

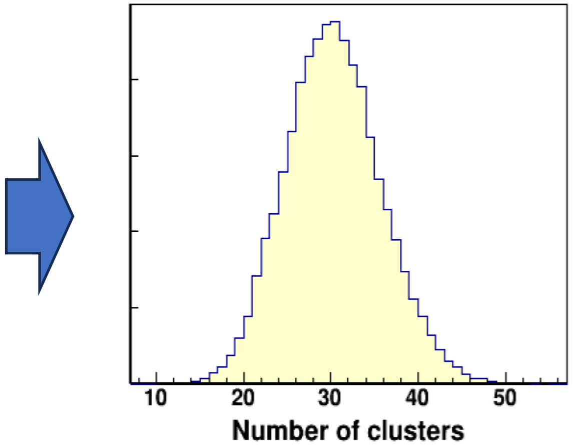
- **Alternatively, counting primary clusters**
  - Poisson distribution ➔ Get rid of the secondary ionizations
  - **Small fluctuation ➔ Potentially, a factor of 2 better resolution than dE/dx**

Waveform

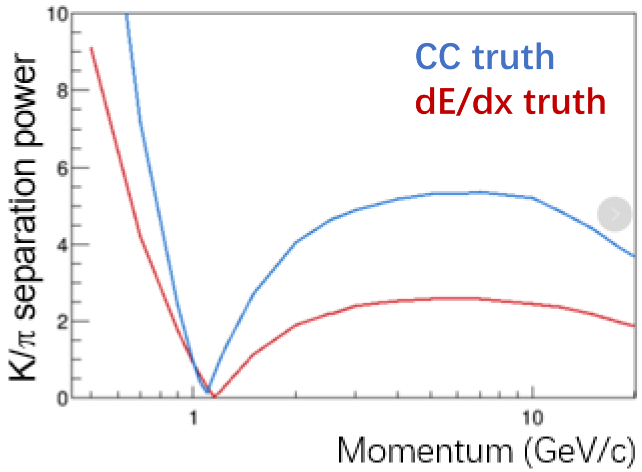


Require fast electronics and sophisticated counting algorithm

Counting clusters



K/π separation power
CC vs dE/dx



CC is extremely powerful, proposed in ILC, FCC-ee, CEPC

Challenges for DC with CC

- **Challenges for cluster counting**
  - Detector design: Detector layout, cell size, working gas with low drift velocity, low ionization density, low diffusion and low cluster size
  - Fast electronics: Bandwidth > 1 GHz, gain > 10, sampling rate > 1.5 GS/s, bit resolution > 12 bit

  → Detector optimization; Prototype experiments

  - Reconstruction: Efficient primary clusters detection from waveforms in high pile-up and noisy environments

  → Deep learning algorithm

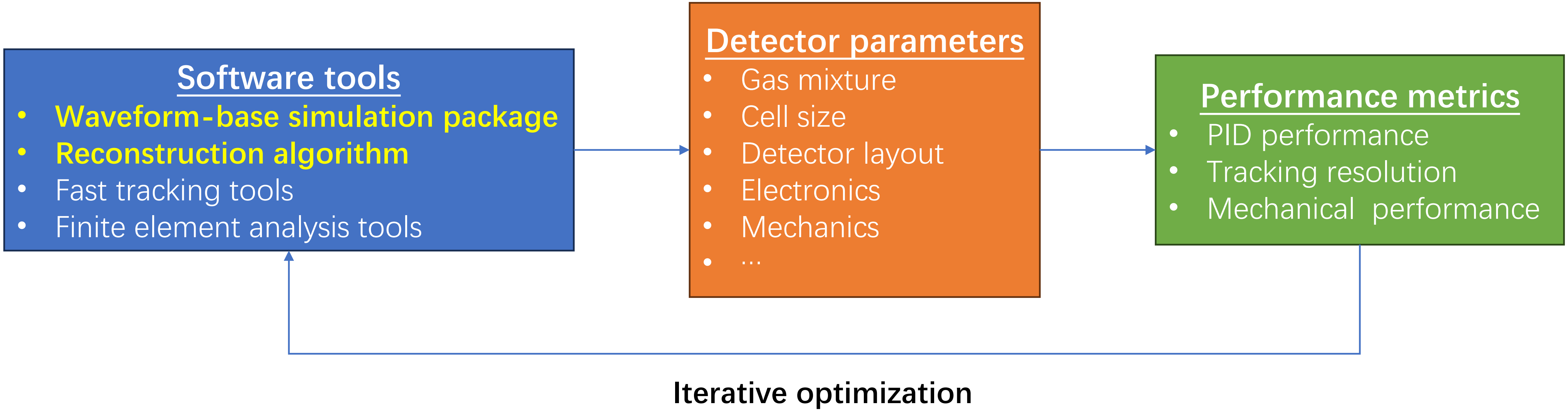
- **Challenges for large volume DC**
  - Electrostatic stability: L ~ 5m, need wire material studies

  → Finite element analysis

  - Data reduction: ~1 TB/s (Z-pole), need online data reduction
  - Power consumption/cooling design

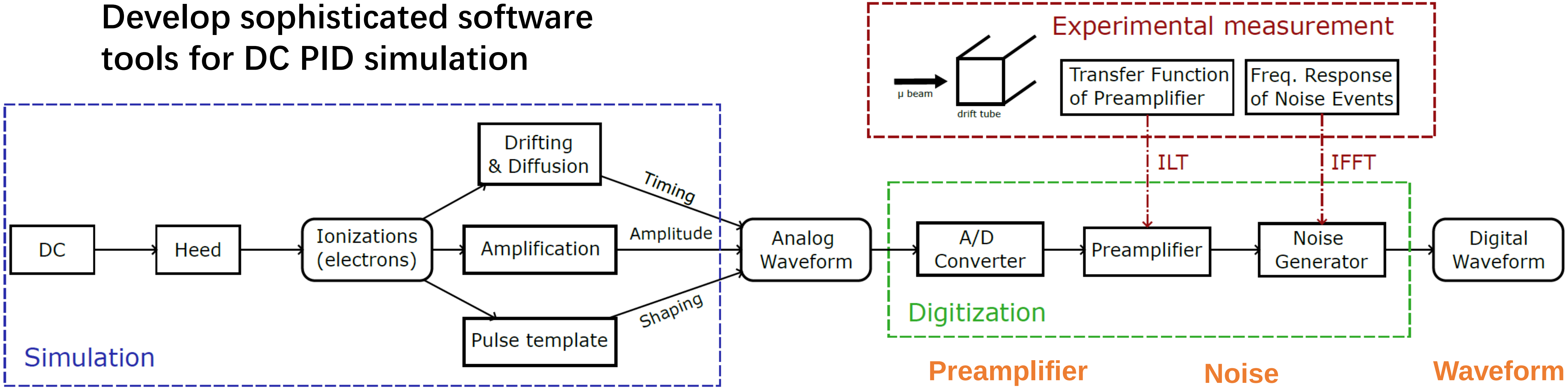
  → Need to do

Overall detector optimization flow

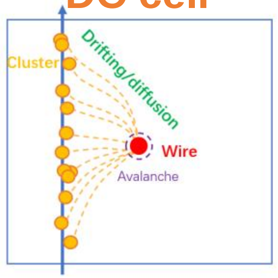


Waveform-based simulation

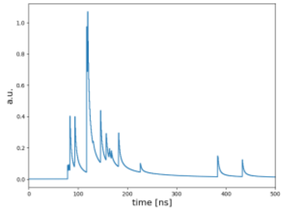
Develop sophisticated software tools for DC PID simulation



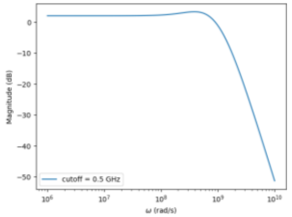
DC cell



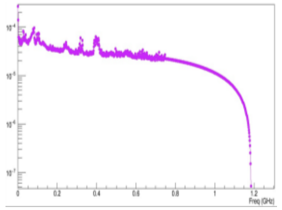
Induced signal



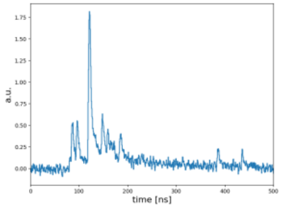
Preamplifier



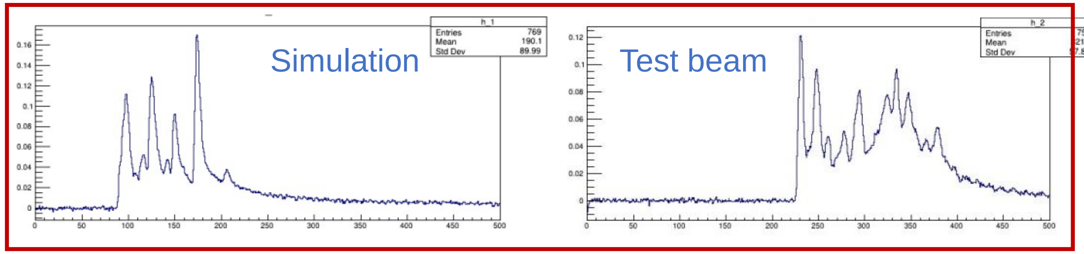
Noise



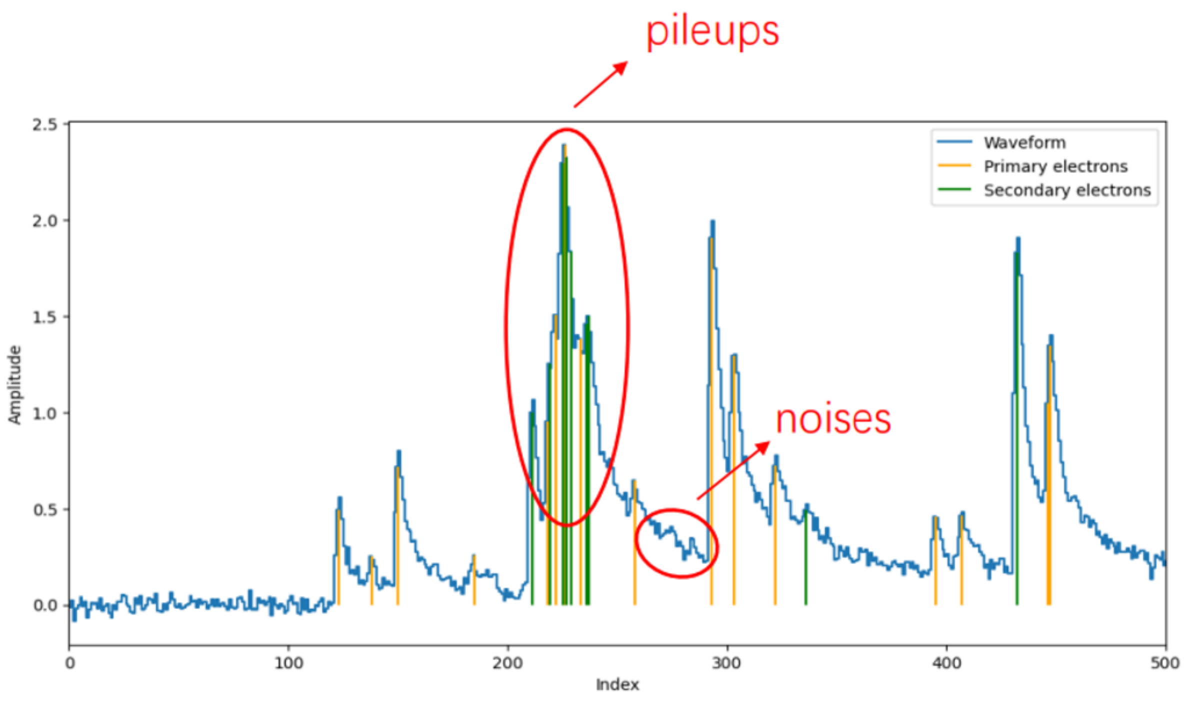
Waveform



Tuned MC is comparable to data



Traditional reconstruction algorithm



Simulated waveform of a DC cell. Orange lines are primary electrons. Green lines are secondary electrons.

Reconstruction: Each primary and secondary electrons forms a peak in the waveform. Need to determine the # of primary peaks.

Peak finding: Detect all electron peaks

- Taking 1st and 2nd order derivatives
- Peak detection by threshold passing

Clusterization: Merge electrons to form clusters

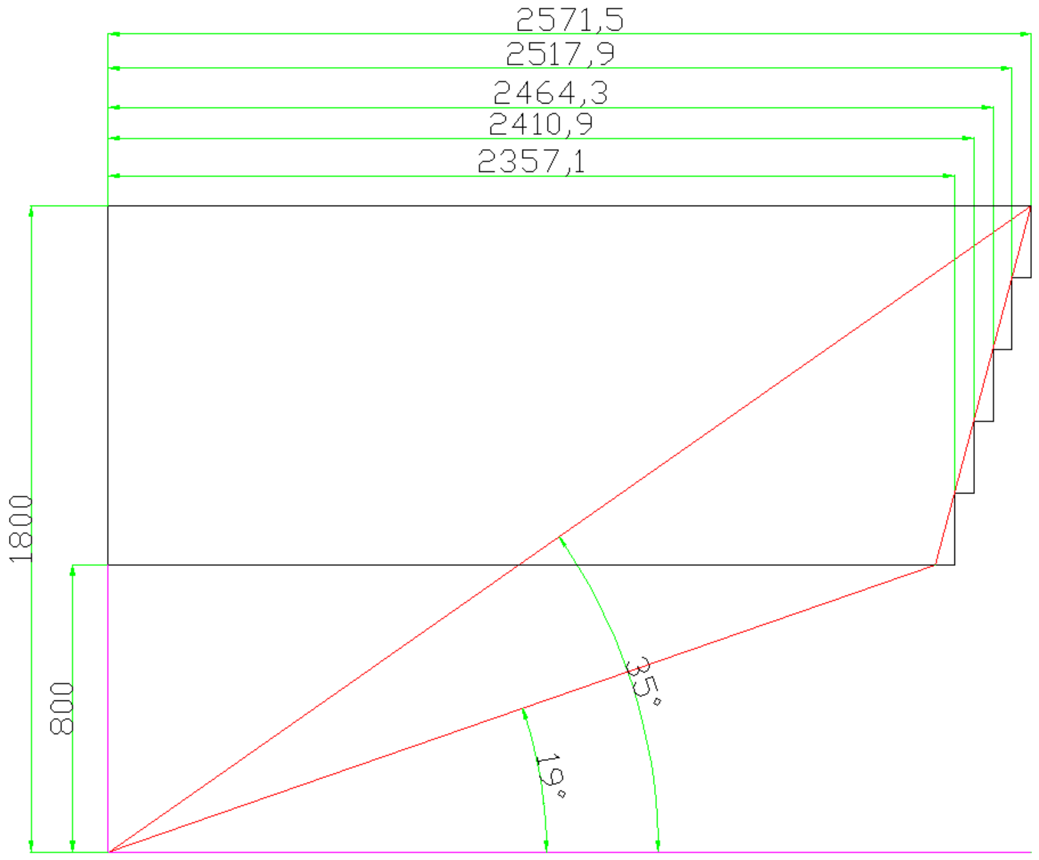
- Merge peaks within $[0, t_{cut})$
- The t_{cut} is related to diffusion

- **Pros:** Fast and easy to implement
- **Cons:** Suboptimal efficiency for highly pile-up and noisy waveforms ➡ **Deep learning**

Preliminary DC design

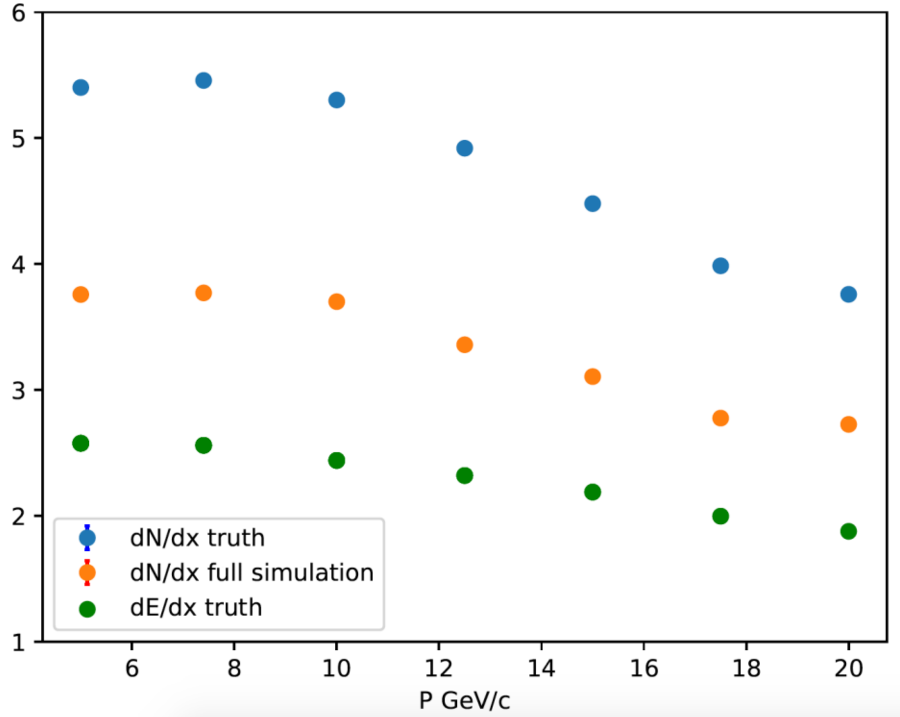
Optimized DC Parameters

DC Parameters	
Radius extension	800-1800 mm
Length of outermost wires（cosθ=0.82）	5143 mm
Thickness of inner CF cylinder	200 μm
Outer CF frame structure	Equivalent CF thickness: 1.63 mm
Thickness of end Al plate	35 mm
Cell size	18 mm × 18 mm
# of cells	24766
Ratio of field wires to sense wires	3:1
Gas mixture	He/iC_4H_{10}=90:10

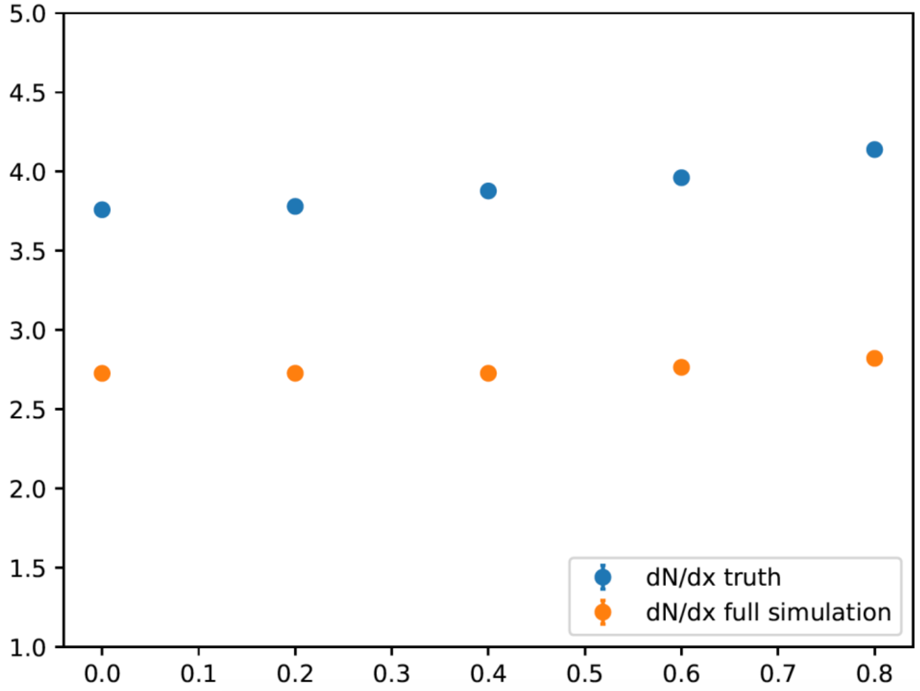


PID performance

K/π separation power vs P (1m track length, cosθ=0)



K/π separation power vs cosθ (P=20GeV/c)

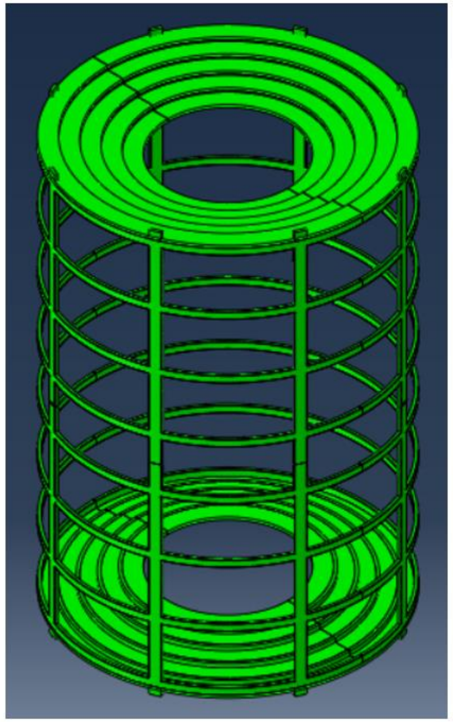


Separation power

$$S = \frac{\left| \left(\frac{dN}{dx}\right)_{\pi} - \left(\frac{dN}{dx}\right)_{K} \right|}{(\sigma_{\pi} + \sigma_{K})/2}$$

2σ K/π separation for 20 GeV/c tracks could be achieved

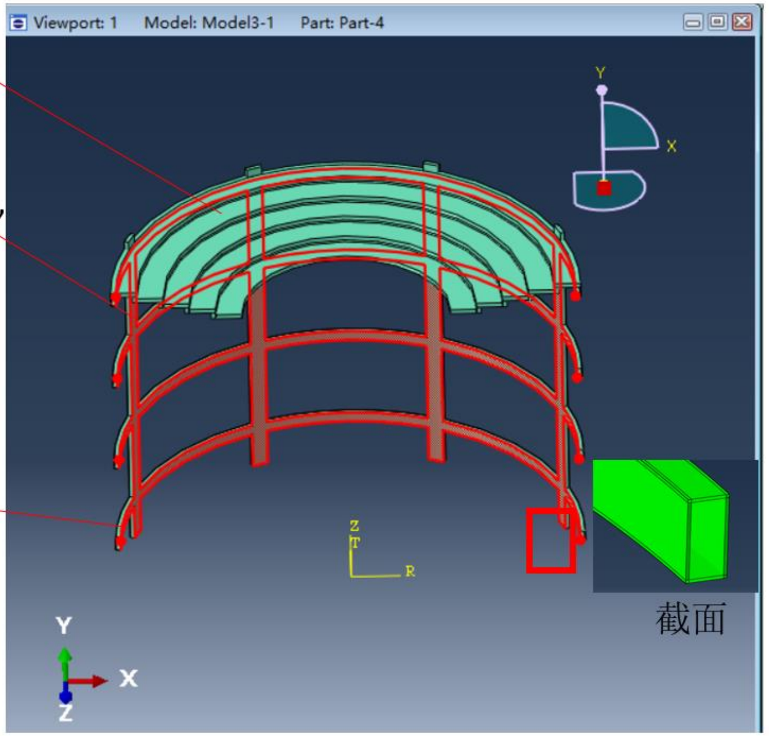
Mechanics with support structures



Al endplates : 35 mm

Longitudinal beam, Cross section: 122×40mm

Annular beam, Cross section: 80×40mm



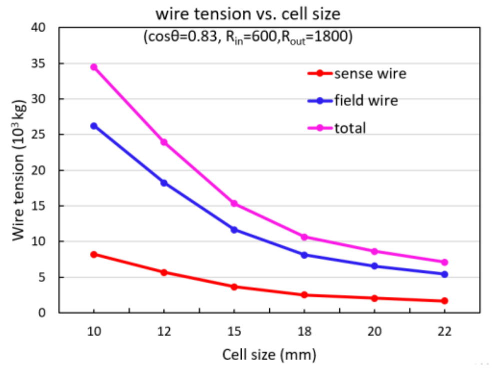
- Carbon fiber frame structure, including 8 longitudinal hollow beams and 8 annular hollow beams
- Thickness of inner CF cylinder: 200 μm/layer
- Effective outer CF frame structure: 1.63 mm
- Thickness of end Al plate: 35 mm

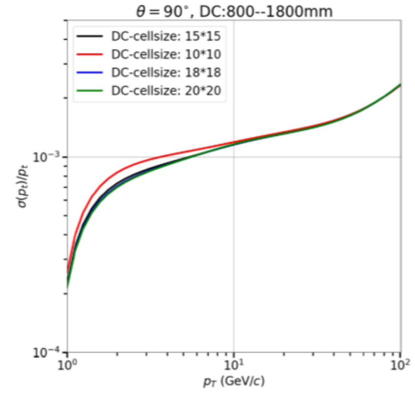
- **Mises stress: 70 MPa**
- **Principal stress: 33 Mpa**
- **Deformation: 0.8 mm**
- **Buckling coefficient: 17.2**

Preliminary calculation shows stable

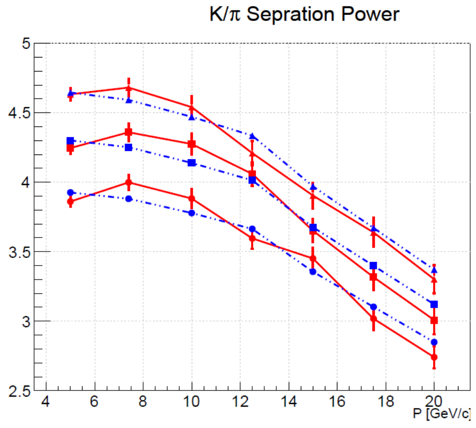
Further optimization considerations

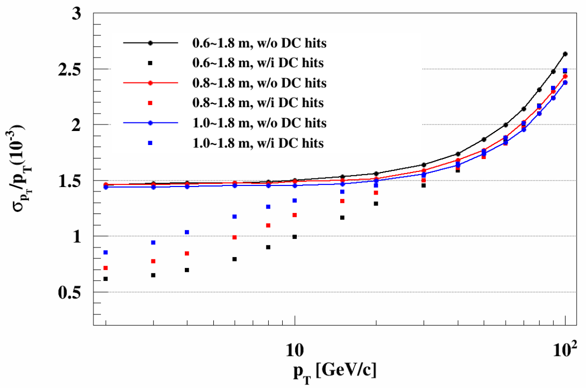
- **Cell size**
  - Cell size has little effect on PID. Large cell size is better for engineering and can reduce # of readout channels
  - Hybrid-cell-size: Reduce the cell size of the first 10 layers to achieve stable operation at high counting rates and minimize aging effects





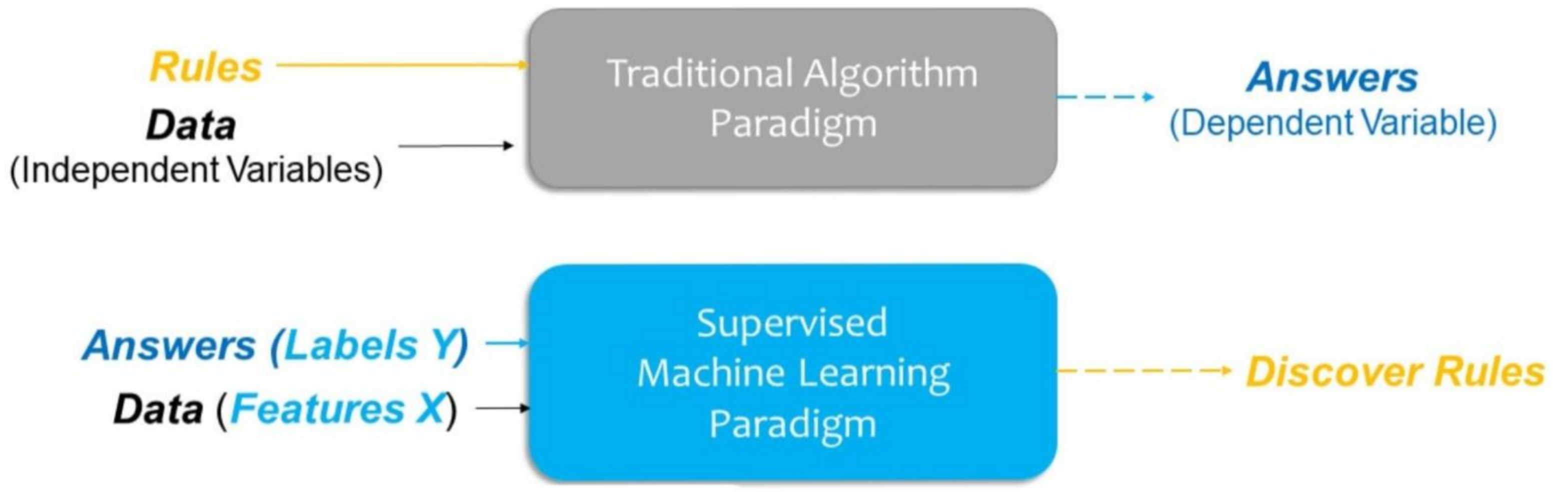
- **Inner radius**
  - Inner radius: 800 mm → 600 mm → 400 mm
  - K/π separation: 2.8σ → 3.1σ → 3.3σ @ 20 GeV/c
  - Smaller inner radius is better for PID and tracking resolution, but more challenge on engineering and aging effects





Reconstruction algorithm with deep learning

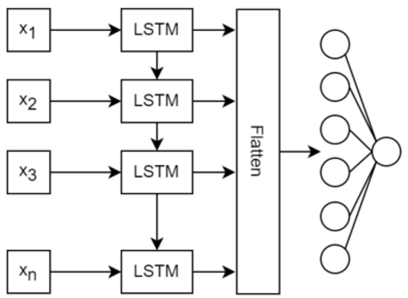
- **Traditional algorithm:**
  - Use partial information of the raw waveform
  - Require human input prior knowledge
- **Supervised learning could be more powerful** because
  - make full use of the waveform information
  - automatically learn characteristics of signals and noises from large labeled samples



Supervised model for simulated samples

Peak finding with LSTM

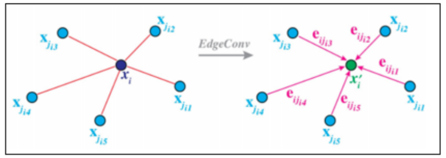
Why LSTM? ➔ Waveforms are time series



- Architecture: LSTM (RNN-based)
- Method: Binary classification of signals and noises on slide windows of peak candidates

Clusterization with DGCNN

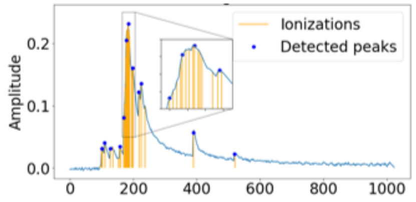
Why DGCNN? ➔ Locality of the electrons from the same primary cluster, perform massage passing through neighbor nodes in GNN



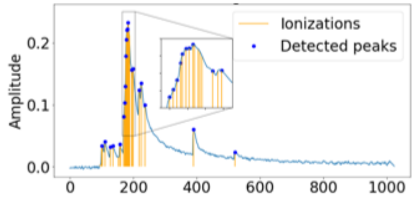
arXiv: 1801.07829

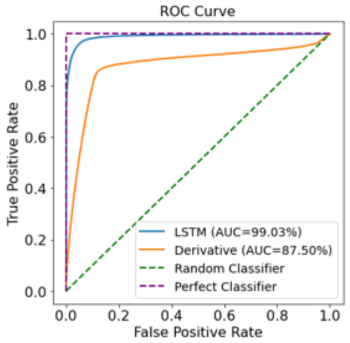
- Architecture: DGCNN (GNN-based)
- Method: Binary classification of primary and secondary electrons

Derivative-based method

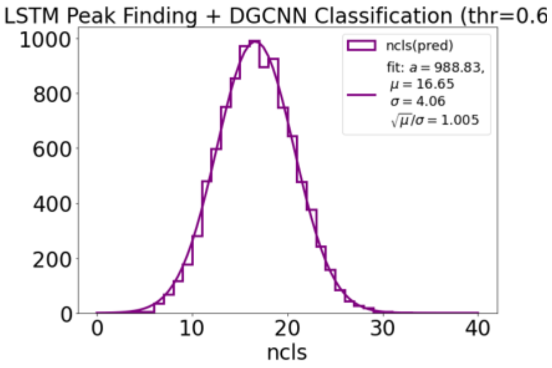


LSTM

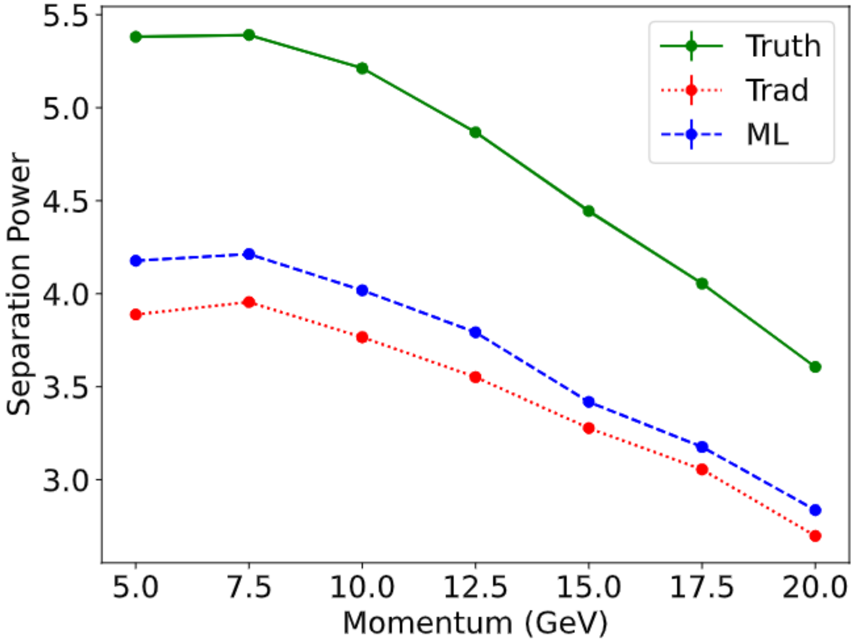




Peak finding: ML is better than derivative-based method



Peak finding + Clusterization: Very well Poisson-like distribution



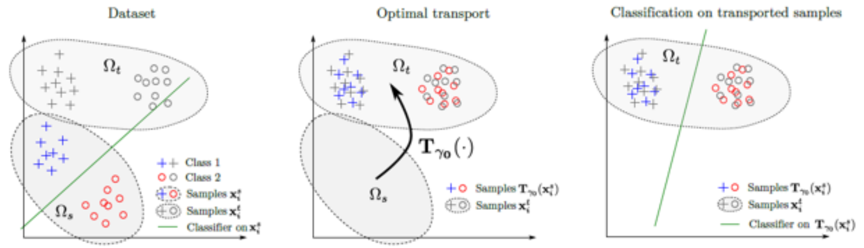
~10% improvement on K/π separation power with ML (equivalent to a detector with 20% larger radius)

Domain adaptation model for data samples

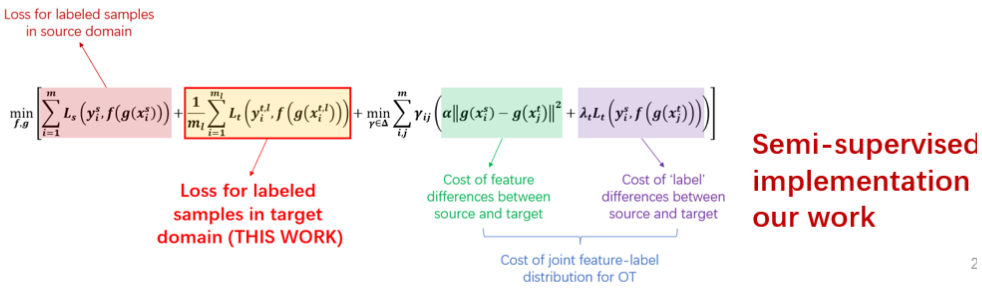
Main challenges:

- Discrepancies between data and MC
- Lack of labels in experimental data

➡ **Domain adaptation**



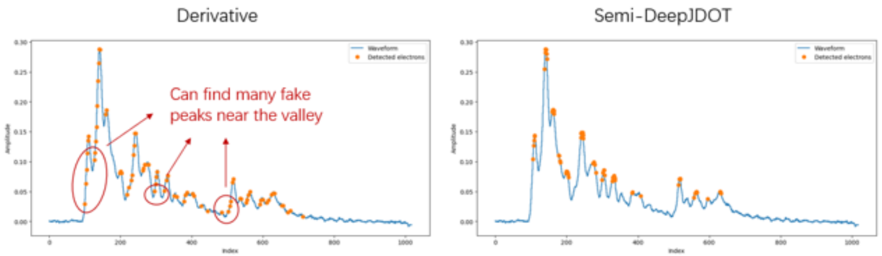
Align data/MC samples with Optimal Transport



$$\min_{f,g}\left[\sum_{i=1}^{m} L_s\left(y_i^s, f\left(g(x_i^s)\right)\right) + \frac{1}{m_l}\sum_{i=1}^{m_l} L_t\left(y_i^{t,l}, f\left(g(x_i^{t,l})\right)\right) + \min_{\gamma\in\Delta}\sum_{i,j}^{m} \gamma_{ij}\left(\alpha\left\|g(x_i^s) - g(x_j^t)\right\|^2 + \lambda_t L_t\left(y_i^s, f\left(g(x_j^t)\right)\right)\right)\right]$$

Semi-supervised implementation our work

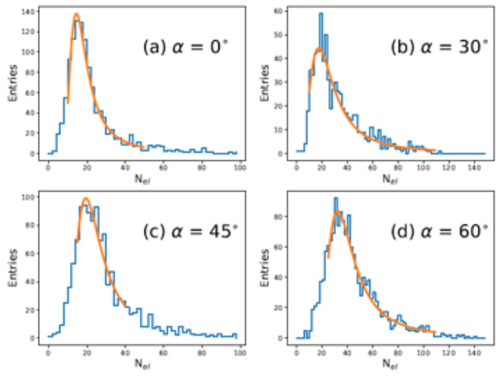
Single-waveform results between derivative alg. and DL alg.

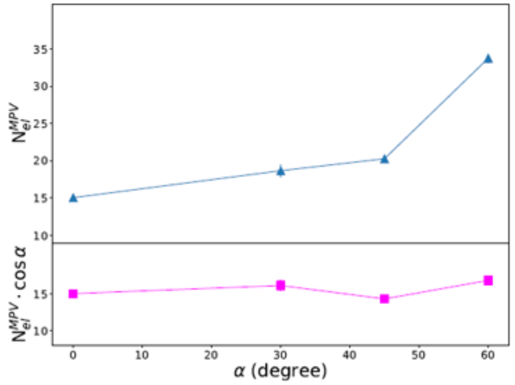


Note: Require similar efficiency for both cases

DL algorithm is more powerful to discriminate signals and noises

Multi-waveform results for samples in different angles





Scale w.r.t. track length

The algorithm is stable w.r.t. track length

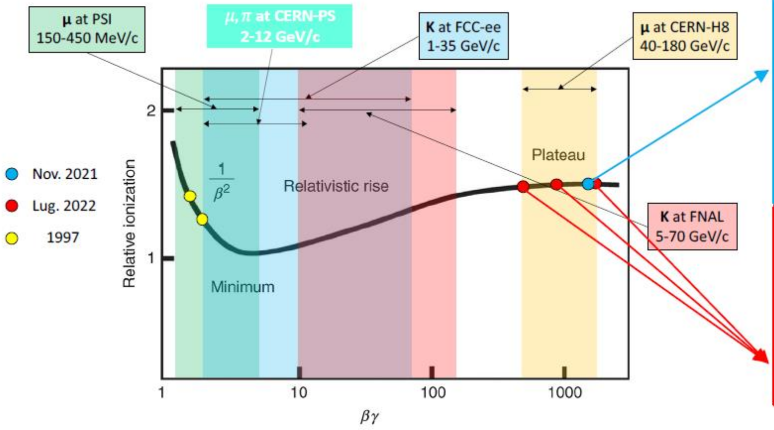
Test beam experiments at CERN

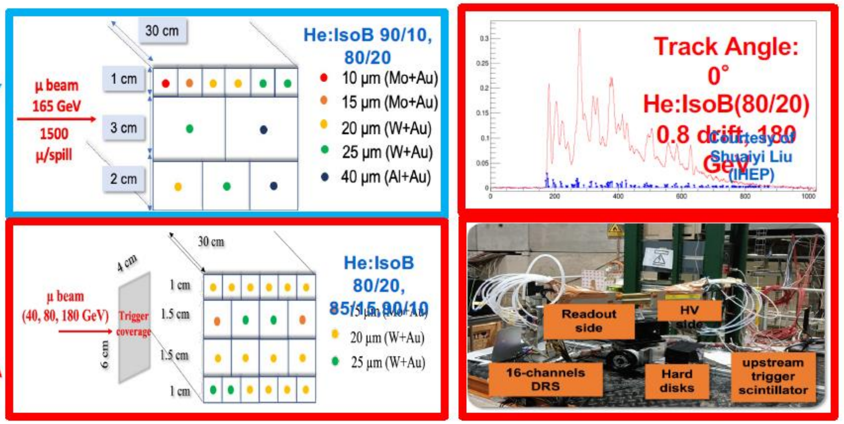
Beam tests organized by INFN group:

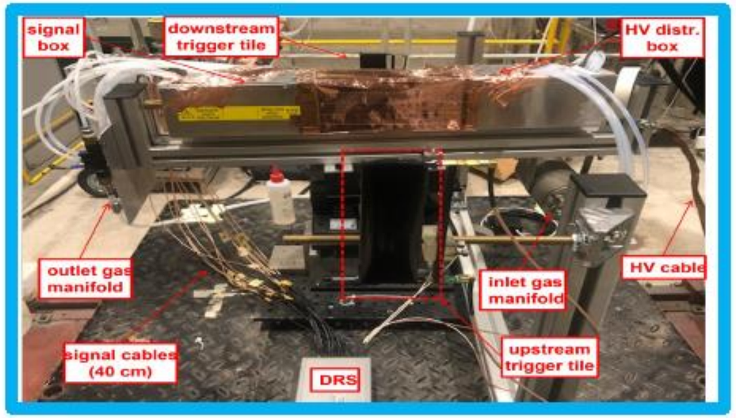
- Two muon beam tests performed at CERN-H8 (βγ>400) in Nov. 2021 and July 2022.
- A muon beam test (from 4 to 12 GeV/c) in 2023 performed at CERN.
- Ultimate test at FNAL-MT6 in 2024 with π and K (βγ = 10-14) to fully exploit the relativistic rise.

Contributions from IHEP group:

- Participate data taking and collaboratively analyze the test beam data
- **Develop the deep learning reconstruction algorithm**



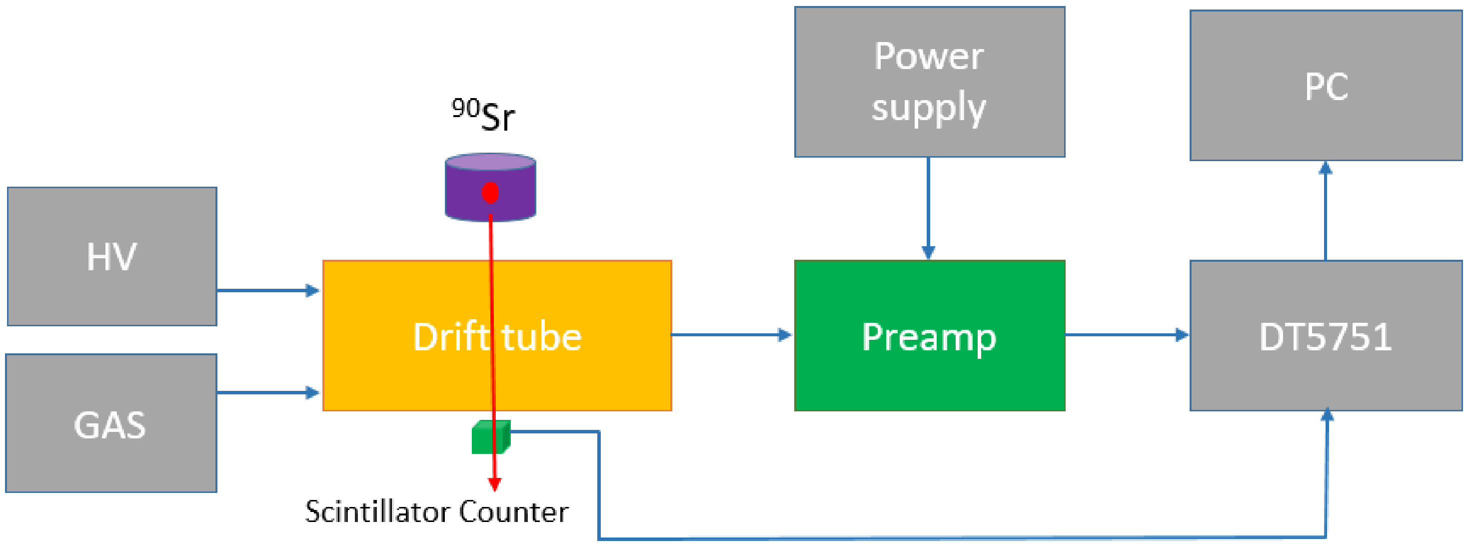




[See Nicola De Filippis's talk at the CEPC Workshop for details]

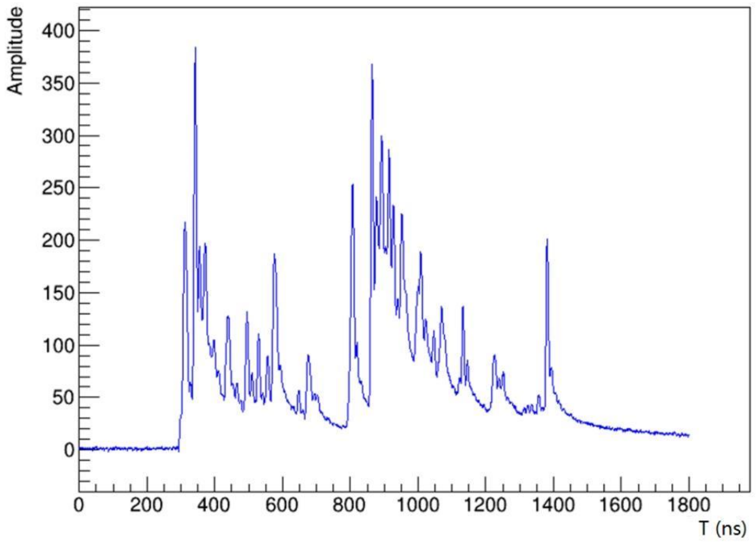
Prototype experiment at IHEP

Experiment layout



- Diameter of the tube: 30 mm
- Working gas: He/iC_4H_{10}=90:10

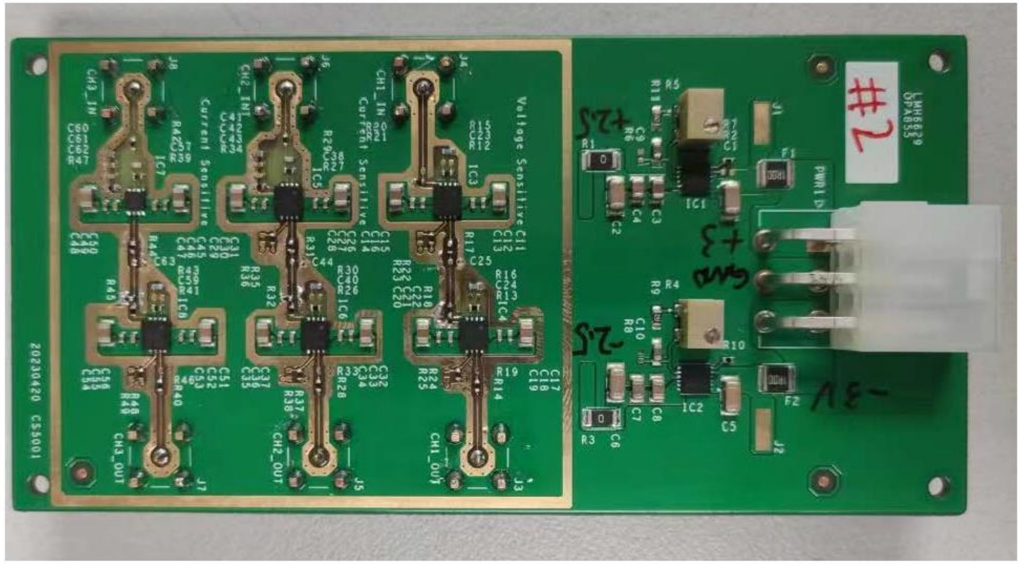
Waveform with Sr-90 β source

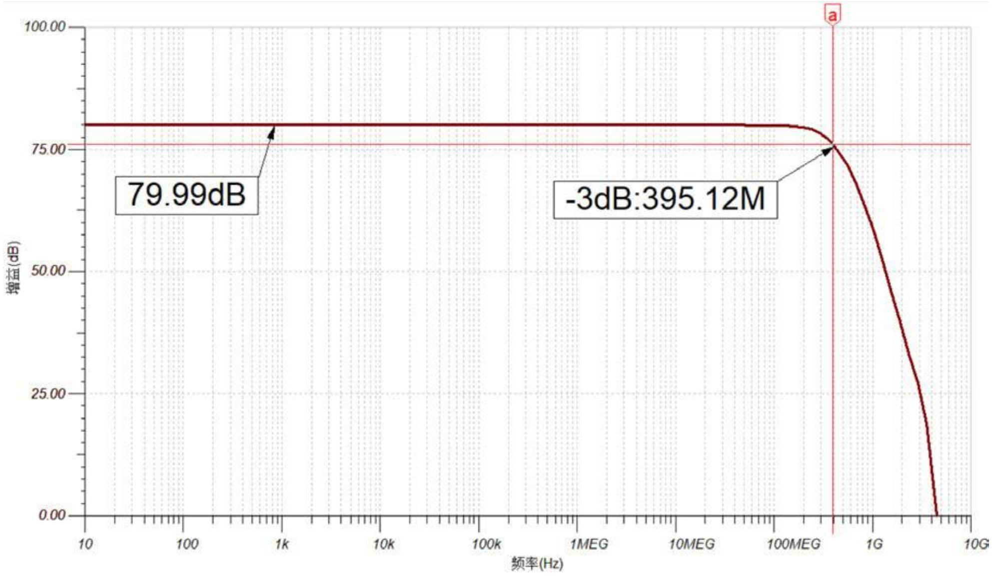


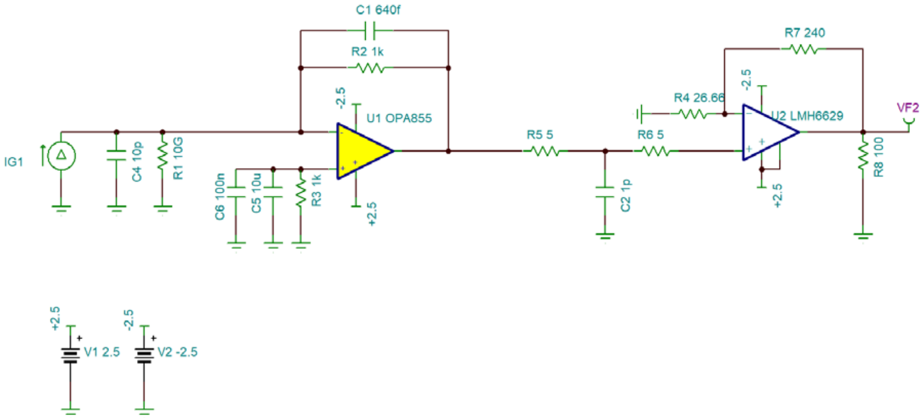
CC signal observed with

- **low noise**
- **high bandwidth**
- **fast risetime: ~ns**

Electronics development







- High bandwidth current sensitive preamplifiers based on LMH6629 have been designed and developed
- Tested with detector prototype and digitizer (DT5751) with 1 GHz sampling rate

Summary

- **Preliminary DC design**
  - PID performance: Close to 3σ K/π separation at 20 GeV/c for 1m track length
  - Mechanical stability: Stable with FEM simulations
- **Reconstruction**
  - 10% improvement on K/π separation for supervised model
  - Domain adaption model for experimental data
- **Experiments**
  - Fast electronics development and prototype test at IHEP: observe CC signals
  - International collaboration on test beam experiments
- **Plans**
  - Fine detector optimization
  - Optimize deep learning algorithm and FPGA implementation
  - Electronics developments and experiments
  - Mechanical design and tests
  - Physics benchmarks

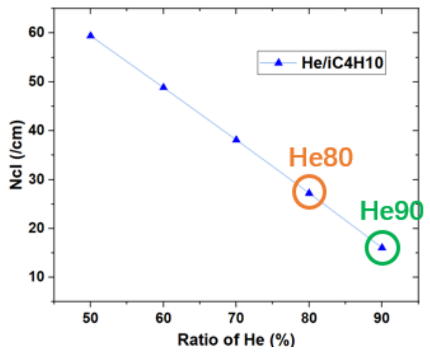
Backup

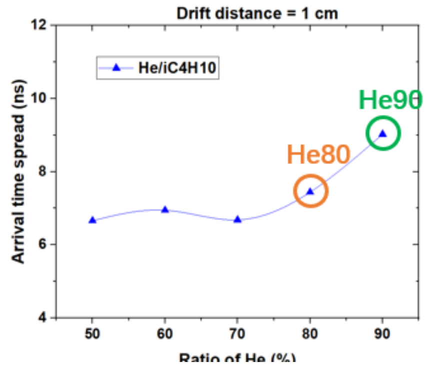
Requirements of detector and key technologies

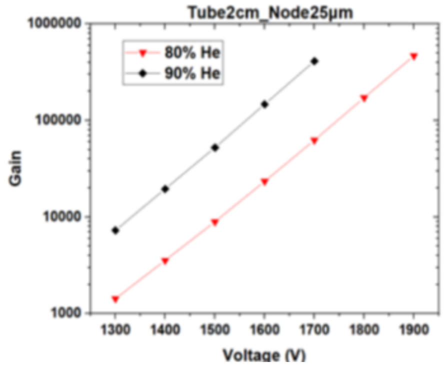
Sub-detector	Key technology	Key Specifications
Silicon vertex detector	Spatial resolution and materials	$\sigma_{r\phi} \sim 3\ \mu\text{m}$, $X/X_0 < 0.15\%$ (per layer)
Silicon tracker	Large-area silicon detector	$\sigma(\frac{1}{p_T}) \sim 2 \times 10^{-5} \oplus \frac{1\times10^{-3}}{p\times\sin^{3/2}\theta}(\text{GeV}^{-1})$
TPC/Drift Chamber	Precise dE/dx (dN/dx) measurement	Relative uncertainty 2%
Time of Flight detector	Large-area silicon timing detector	$\sigma(t) \sim 30$ ps
Electromagnetic Calorimeter	High granularity 4D crystal calorimeter	EM energy resolution $\sim 3\%/\sqrt{E(\text{GeV})}$ Granularity $\sim 2 \times 2 \times 2\ \text{cm}^3$
Magnet system	Ultra-thin High temperature Superconducting magnet	Magnet field $2-3$ T Material budget $< 1.5X_0$ Thickness < 150 mm
Hadron calorimeter	Scintillating glass Hadron calorimeter	Support PFA jet reconstruction Single hadron $\sigma_E^{had} \sim 40\%/\sqrt{E(\text{GeV})}$ Jet $\sigma_E^{jet} \sim 30\%/\sqrt{E(\text{GeV})}$

Gas mixture

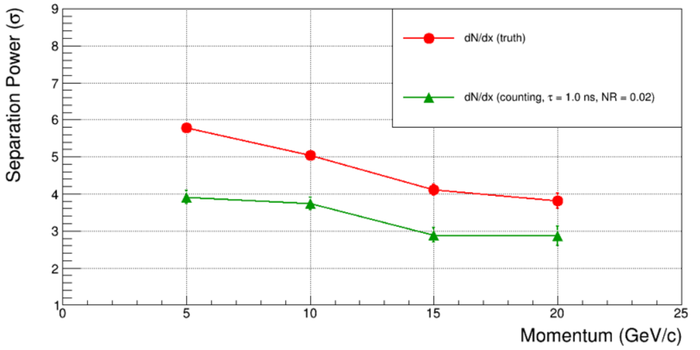
- **Gas mixture choice: He + C_4H_{10}**



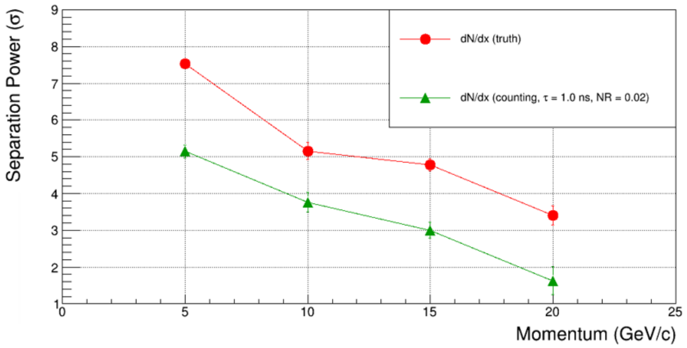




He 90% + iC_4H_{10} 10%, L = 1m, NR = 0.02，1x1cm cell



He 80% + iC_4H_{10} 20%, L = 1m, NR = 0.02, 1x1cm cell

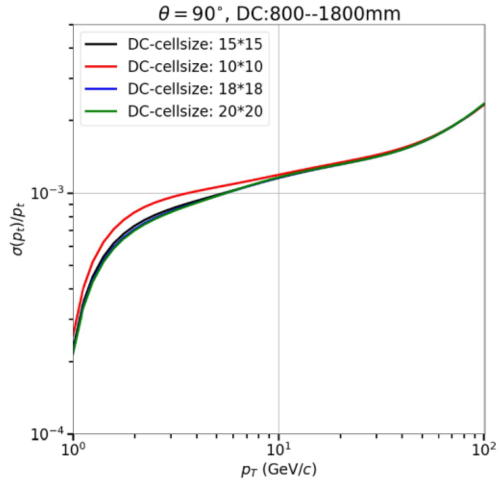


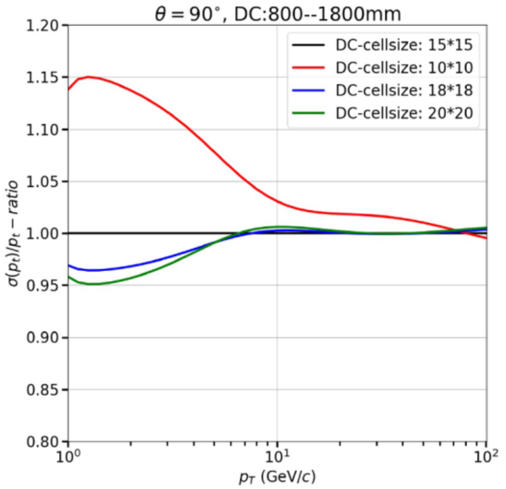
Gas property for CC:

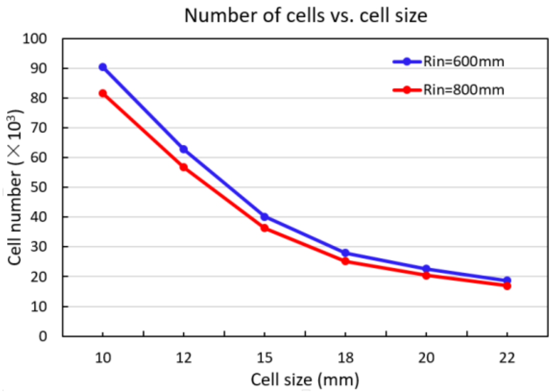
- Small ρ_{cl} ➔ less statistics, large time separation
- Slow v_d ➔ large time separation
- Small σ_d ➔ less likely double-counting

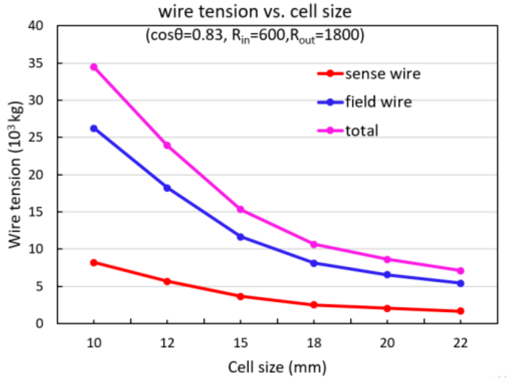
- **He 90% + iC_4H_{10} 10% is better for high momentum**

Cell size

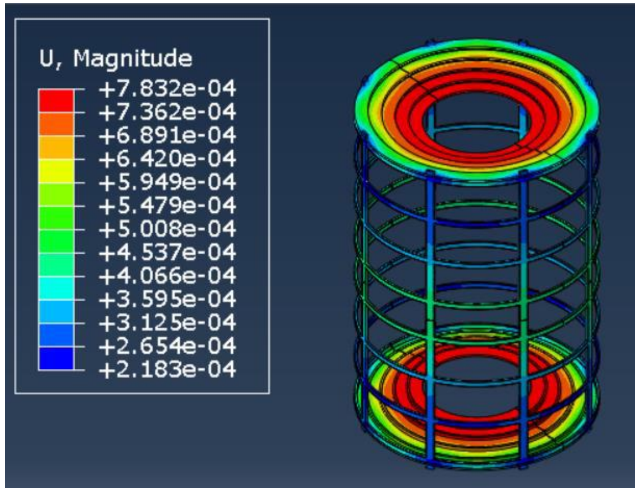








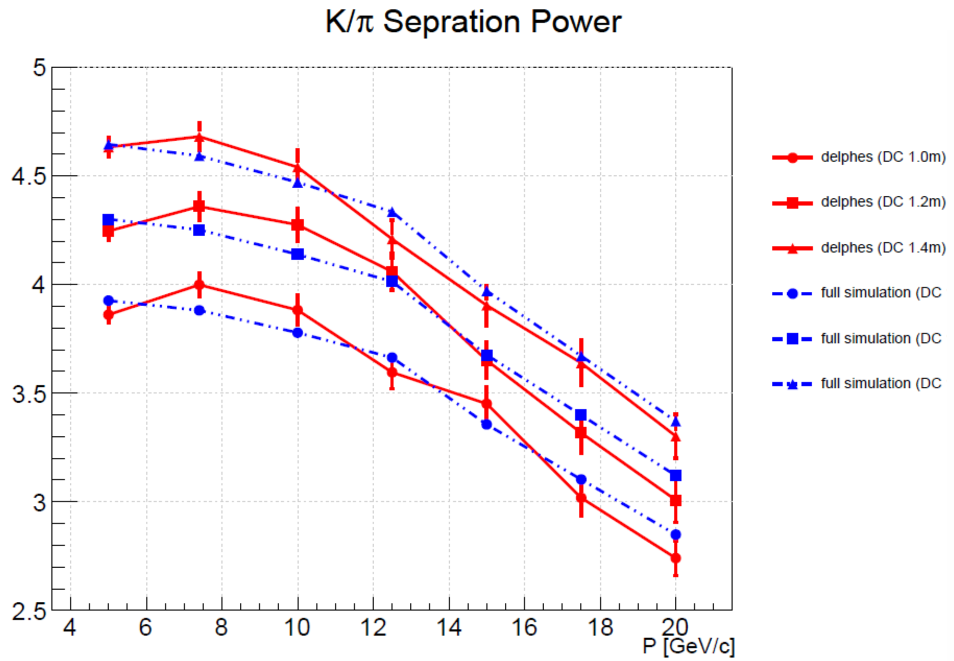
- Increasing the cell size has little effect on the PID performance.
- The number of wires can be reduced, hence the production difficulty, the number of readout channels, and the material of the supporting structure (mostly at the outer cylinder).
- Hybrid-cell-size could be possible. (e.g. reduce the size of the first 10 layers)

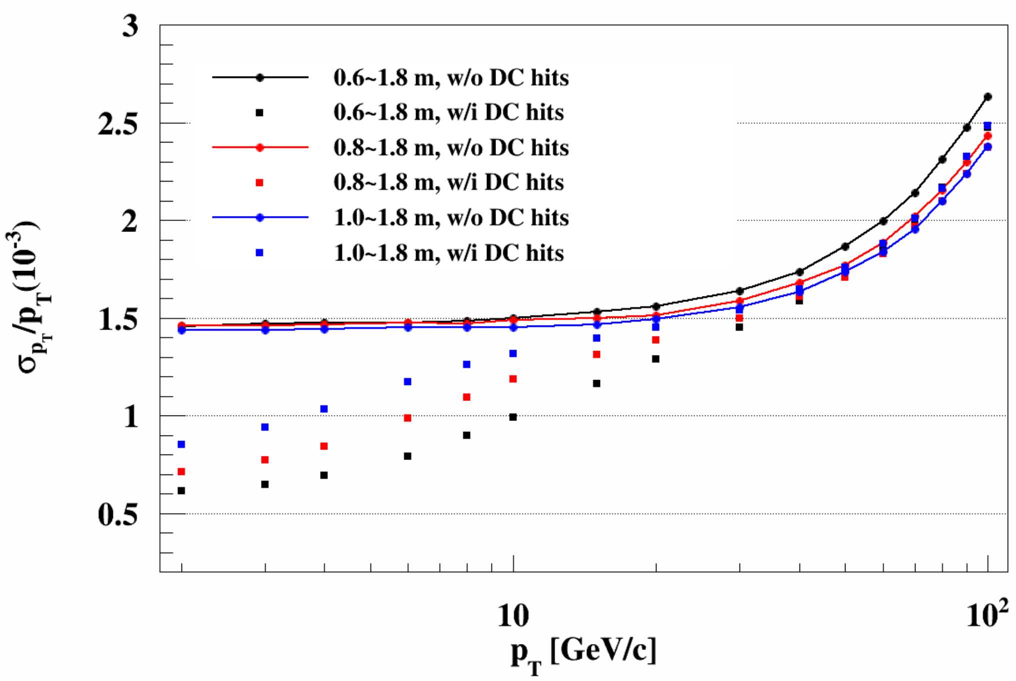


Finite element analysis

With cell size of 18x18 mm^2, field (sense) wire with Φ=60 (20) μm, wire tension ~ 9000 kg which satisfies stability condition

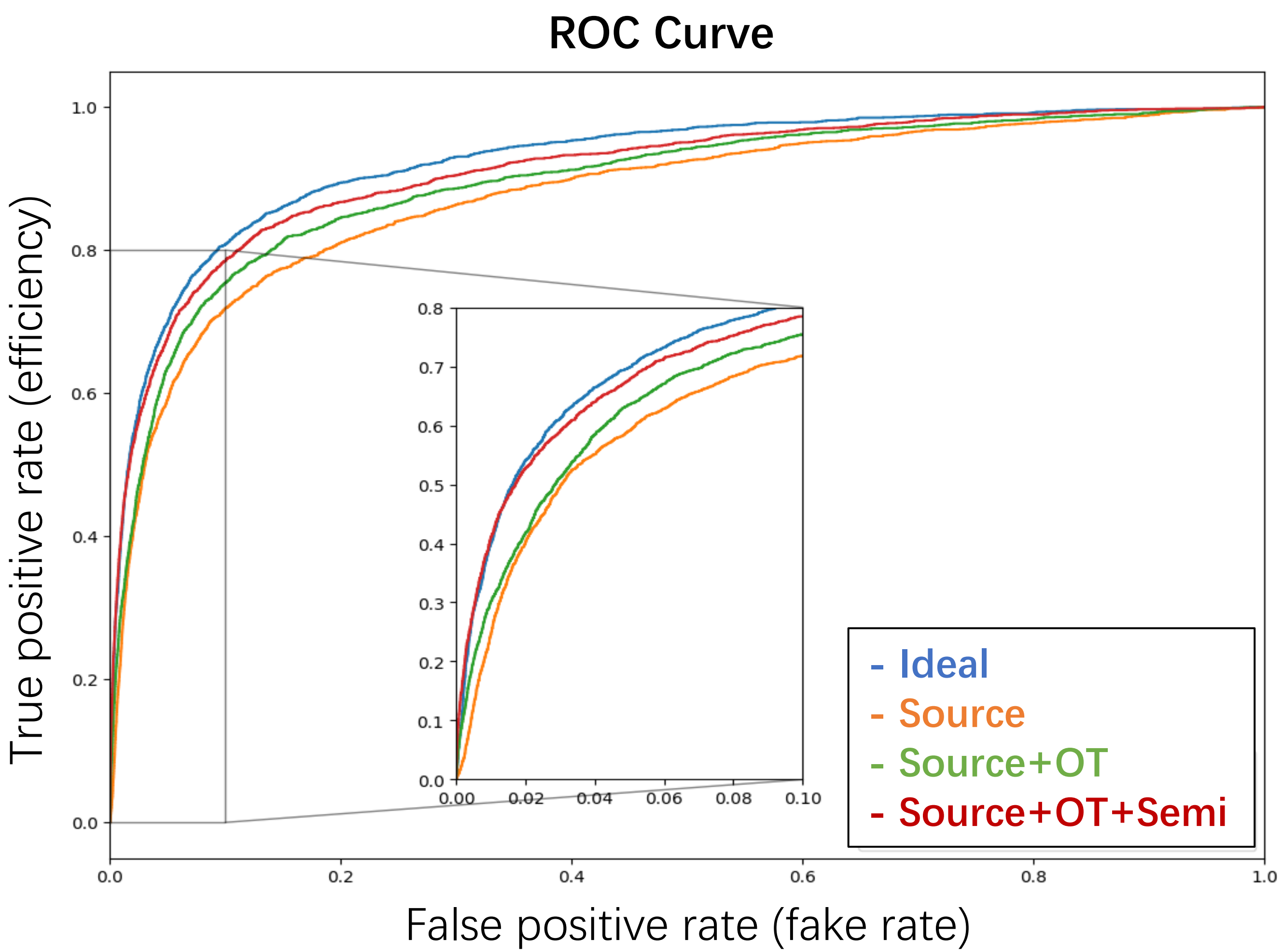
Inner radius





- Large thickness is better for PID ➔ From 2.8σ to 3.1σ even 3.3σ @ 20 GeV/c
- Smaller inner radius is better for tracking resolution, but more challenge on engineering and more beam backgrounds

Numeric experiment



Numeric experiment with pseudo data:

- Use labels in pseudo data to evaluate

Model	AUC	pAUC (FPR<0.1)
Ideal (supervised)	0.926	0.812
Source (baseline)	0.878	0.749
Source + OT	0.895	0.769
Source + OT + Semi (Semi-supervised DA)	0.912	0.793

Validation: Performance of Semi-DeepJDOT model is very close to the ideal model (supervised model)

Mechanicals: Wire tension

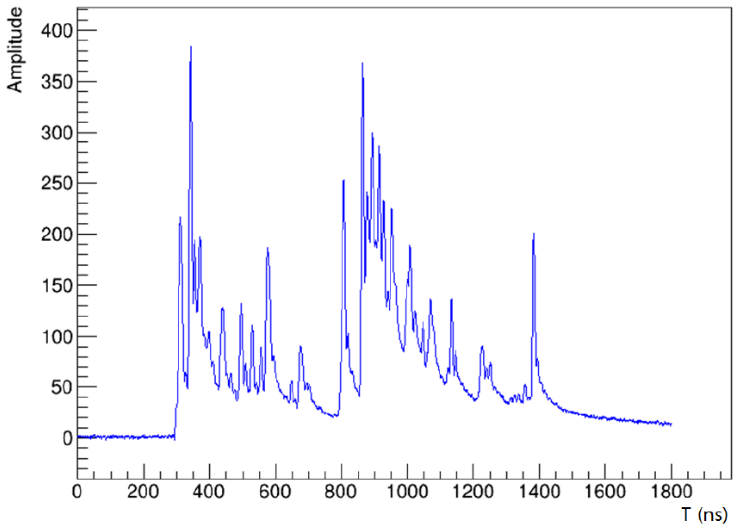
- ✓ Diameter of field wire (Al coated with Au) : 60 µm
- ✓ Diameter of sense wire (W coated with Au): 20 µm
- ✓ Sag = 280 µm

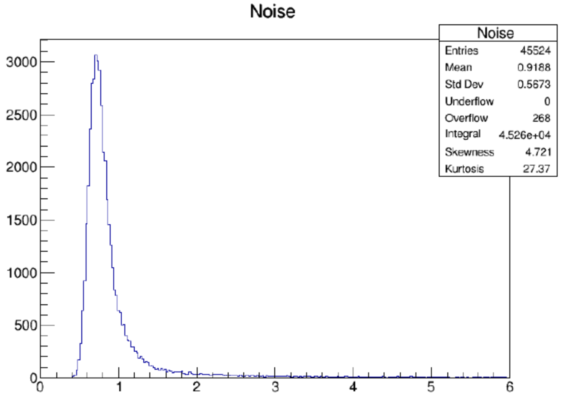
Step	cell number /step	length	single sense wire tension (g)	Single field wire tension (g)	total tension /step (kg)
1	3417	4715	60.15	92.42	1153.08
2	4185	4822	62.91	96.66	1477.02
3	4953	4929	65.74	101.00	1826.47
4	5721	5036	68.62	105.44	2202.24
5	6489	5143	71.57	109.96	2605.11
total	**24766**				**9263.92**

Meet requirements of stability condition: $T > (\frac{VLC}{d})^2/(4\pi\varepsilon_0)$

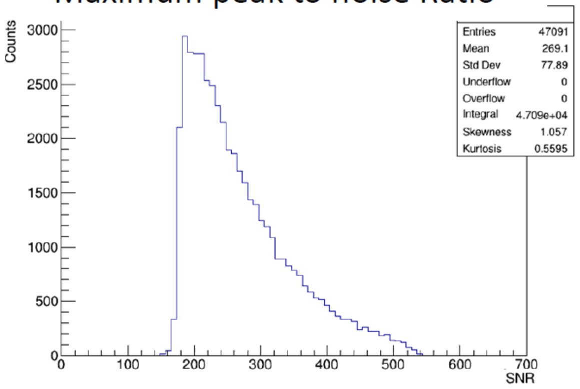
Waveforms from prototype tests

Waveform with Sr-90 β source

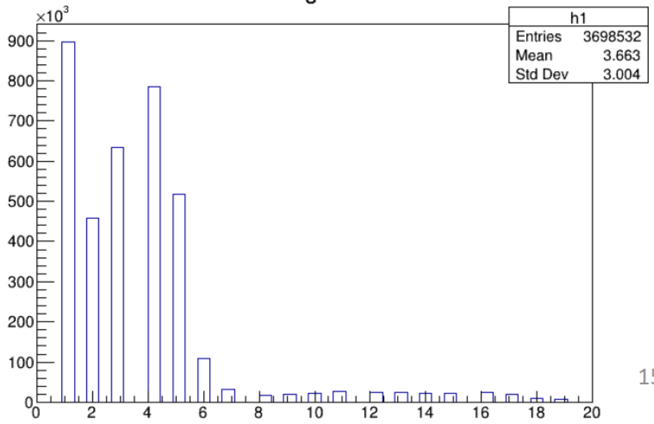




Maximum peak to noise Ratio



Rising Time

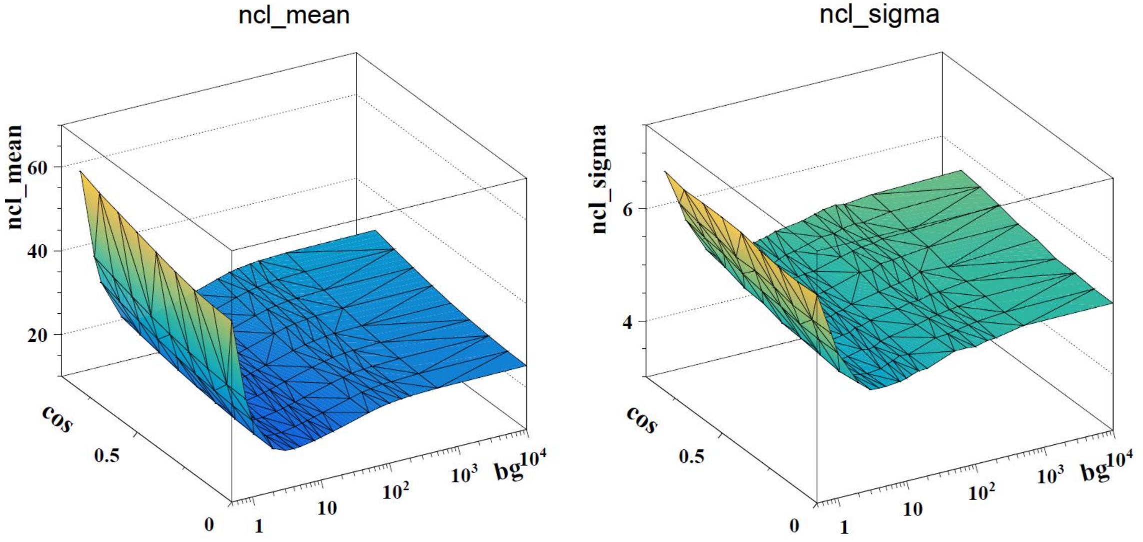


15

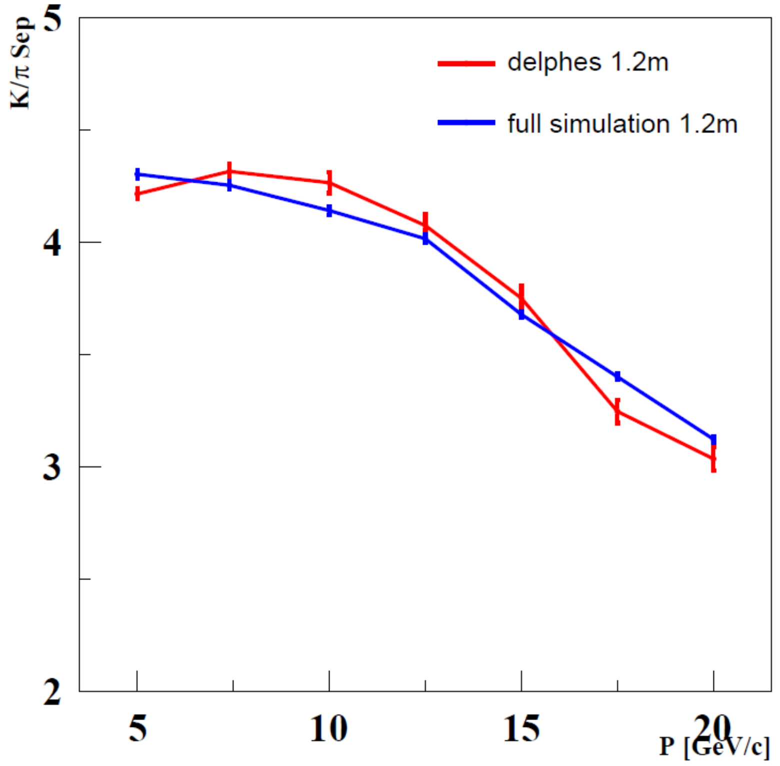
- Test results show
  - low noise
  - high bandwidth
  - Rise time of peak: a few nanoseconds

Physics study with Delphes

- **Delphes: A C++ framework, performing a fast multipurpose detector response simulation**
  - $10^2 \sim 10^3$ faster than the fully GEANT-based simulations
  - Sufficient and widely used for phenomenological studies
- **Develop dedicated PID modules (CC and TOF) and perform quick physics studies**



K/π separation power

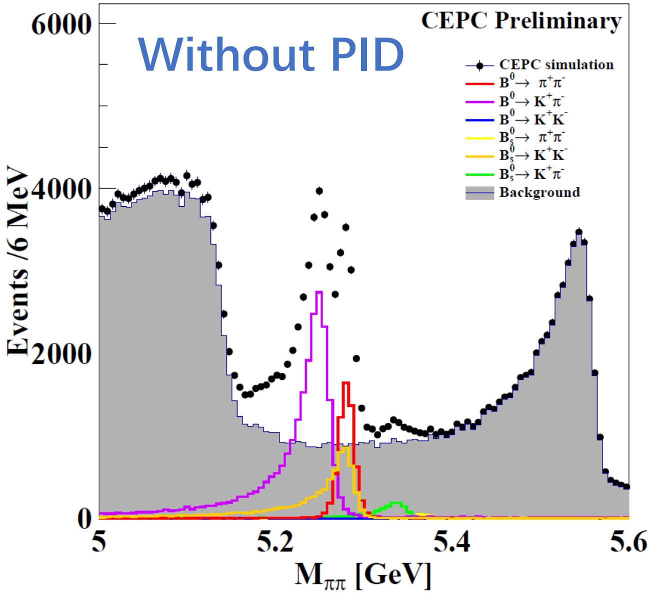


Good consistent to full simulation

Study of $\mathbf{B^0_{(s)} \to h^+h'^-}$

- **Motivation**
  - Rich physics programs in $B^0_{(s)} \to h^+h'^-$ decays
    - Time-dependent asymmetry, direct CP violation, lifetime measurement, ⋯
  - Good test bed to study impact of PID in flavor physics
  - Explore physics potential of Tera-Z

- **Significantly improved SNR with PID**
- **More detailed studies ongoing**



With PID